

A linear regression pattern for electricity price forecasting in the Iberian electricity market

Patrón de regresión lineal para la previsión de precios de electricidad del mercado eléctrico Ibérico

Ângela Paula Ferreira ^{1, 2*}, Jenice Gonçalves Ramos ¹, Paula Odete Fernandes ^{1, 3}

¹Polytechnic Institute of Bragança, Campus de Santa Apolónia.C.P. 5300-253. Bragança, Portugal.

²Research Centre in Digitalization and Intelligent Robotics (CeDRI), Polytechnic Institute of Bragança, Campus de Santa Apolónia. C.P. 5300-253. Bragança, Portugal.

³Applied Management Research Unit (UNIAG), Polytechnic Institute of Bragança, Campus de Santa Apolónia. C. P. 5300-253. Bragança, Portugal.

CITE THIS ARTICLE AS:

Â. P. Ferreira, J. Gonçalves and P. O. Fernandes. "A linear regression pattern for electricity price forecasting in the Iberian electricity market", *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 93, pp. 117-127, Oct-Dec 2019. [Online]. Available: <https://www.doi.org/10.17533/udea.redin.20190522>

ARTICLE INFO:

Received: November 01, 2018

Accepted: May 28, 2019

Available online: May 28, 2019

KEYWORDS:

Forecasting, regression analysis, MIBEL

Previsión, análisis de regresión, MIBEL

ABSTRACT: The Iberian Market for Electricity resulted from a cooperation process developed by the Portuguese and Spanish administrations, aiming to promote the integration of the electrical systems of both countries. This common market consists of organised markets or power exchanges, and non-organised markets where bilateral over-the-counter trading takes place with or without brokers. Within this scenario, electricity price forecasts have become fundamental to the process of decision-making and strategy development by market participants. The unique characteristics of electricity prices such as non-stationarity, non-linearity and high volatility make this task very difficult. For this reason, instead of a simple time forecast, market participants are more interested in a causal forecast that is essential to estimate the uncertainty involved in the price. This work focuses on modelling the impact of various explanatory variables on the electricity price through a multiple linear regression analysis. The quality of the estimated models obtained validates the use of statistical or causal methods, such as the Multiple Linear Regression Model, as a plausible strategy to achieve causal forecasts of electricity prices in medium and long-term electricity price forecasting. From the evaluation of the electricity price forecasting for Portugal and Spain, in the year of 2017, the mean absolute percentage errors (MAPE) were 9.02% and 12.02%, respectively. In 2018, the MAPE, evaluated for 9 months, for Portugal and Spain equals 7.12% and 6.45%, respectively.

RESUMEN: El Mercado Ibérico de Electricidad resulta del proceso de cooperación entre las administraciones de Portugal y España con el objetivo de promover la integración de los sistemas eléctricos de ambos países. Este mercado común consiste en mercados organizados o intercambios de potencia, y mercados no organizados donde el comercio bilateral de venta libre se realiza con o sin intermediarios. Dentro de este escenario, la previsión de los precios de energía ha tomado un papel fundamental en el proceso de decisión y estrategia de desarrollo para los mercados participantes. Las características de precios de energía tales como la no-estacionalidad, no linealidad y la alta volatilidad hace que este trabajo sea más complejo. Así, en lugar de una simple previsión, los participantes están más interesados en las causas, que es esencial para estimar el precio. Esta investigación analiza el impacto de variables externas en los precios de electricidad utilizando un modelo de regresión lineal. La calidad de los modelos estimados obtenidos valida el uso de métodos estadísticos o causales, como una estrategia plausible para obtener previsiones causales de los precios de la electricidad a mediano y largo plazo. A partir de la evaluación de la previsión del precio de la electricidad para Portugal y España, para el año 2017, los errores porcentuales absolutos medios (MAPE) fueron de 9.02% y de 12.02%, respectivamente. Para el año de 2018, el MAPE, evaluado para 9 meses, para Portugal y España equivale a 7.12% y 6.45%, respectivamente.

* Corresponding author: Ângela Paula Ferreira

E-mail: apf@ipb.pt

ISSN 0120-6230

e-ISSN 2422-2844



1. Introduction

The Iberian Market for Electricity (MIBEL) outcomes from a cooperative process developed by the Portuguese and

DOI: 10.17533/udea.redin.20190522

Spanish governments, aiming at promoting the integration of the electrical systems and markets of both countries within a framework for providing access to all interested parties under the terms of equality, transparency and objectivity. Trading within MIBEL is done in a free competitive regime, despite the need to comply with market rules, applicable legislation, competition rules and regulation on wholesale energy market integrity and transparency.

There are two organised markets under the MIBEL framework, apart from the system services' market existing in each country: the spot (day ahead and intraday) market, operated by the Spanish branch of MIBEL (OMIE) and abiding by the Spanish legislation, and the forwards and derivatives market, operated by the Portuguese branch of MIBEL (OMIP) and abiding by the Portuguese legislation.

The OMIE market works as a single market for Portugal and Spain if the available interconnection capacity between both countries is sufficient to perform supply and demand orders. When the interconnection capacity becomes technically insufficient, markets are separated, and specific prices are produced for each market under a market splitting mechanism.

With the MIBEL implementation, the Iberian electricity market was moved to an organised, liberalised market regime, which was also an important step in the consolidation of the European Electricity Market. In this sense, it became possible for any Iberian consumer to buy electricity from any producer or marketer operating in Portugal or Spain, under a regime of free competition [1].

The genuine role of the organized market for electricity is to match the supply and the demand of electricity in order to determine the market clearing price. The market price is established in an auction, conducted in a periodical basis for each of the load periods, as the intersection between the supply curve, constructed from aggregated supply bids, and the demand curve, constructed from aggregated demand bids or the system operator estimated demand [2]. Buy/sell orders are accepted in order of increasing (decreasing) prices until the total demand (supply) is met.

Electricity is a very special commodity, being technically and economically non-storable. Besides, power system stability requires a constant balance between production and consumption, which in turn, depends on climate conditions, the intensity of business and everyday activities. Due to the liberalized nature of the market, electricity prices acquire uncertain and volatile characteristics, which can be up to two orders of magnitude higher than any other commodity or financial

assets [3]. In this competitive environment, it is imperative to predict the future price of electricity, aiming at the definition of a dispatch strategy, investment profitability analysis and planning, increasing the profit of energy producers and assisting a decrease in the electricity price for consumers.

Although the wholesale of electricity reflects the real-time cost for supplying which varies minute by minute, the cost formation of electricity prices for final consumers, investment profitability analysis and planning are based on an average seasonal cost. In this regard, the main objective of this work is the construction of statistical (or casual) models to forecast electricity prices, in a monthly basis, in the time span of 2017 and 2018 years, through the Multiple Linear Regression Model (MRLM). A simplified version of this manuscript was previously published as a conference paper [4]. The research has been extended, including the analysis of four new exogenous variables able to impact in the electricity price forecasting in the Iberian countries.

This manuscript is organised as follows: section 2 presents the main factors that may contribute to the variability of electricity prices; section 3 introduces and discusses the forecasting methodology, while section 4 presents and discusses its application to the Iberian countries. Finally, section 5 draws the main conclusions of the performed analysis.

2. Key factors affecting electricity prices

Unique features of electric energy pricing such as non-stationarity, non-linearity and high volatility make the forecast of electricity prices a difficult task. For this reason, instead of a simple one-off forecast, market players are more interested in a causal forecast able to estimate the uncertainty involved in the price. Therefore, it is necessary to analyse the variables that can explain, even though partially, the variability of prices under a long-term basis forecasting horizon, with lead times measured in months.

A large number of external variables may explain the electricity price dynamics, but there is little evidence on the degree and sign of these influences. Exogenous variables such as generation capacity, load profiles and ambient conditions have been previously used in literature to explain the electricity price dynamics. For instance, power consumption, water supply air temperature and load profiles were used in [5–7]. The forecast of zonal electricity prices in Italy, as performed in [8], explored the

effect of technologies, market power, network congestions and demand.

This work analyses several exogenous variables, exploiting the demand, ambient conditions, production of goods, energy sources (renewable and non-renewable) and the import and export energy balance.

Demand and ambient conditions pose a significant contribution to the electricity price dynamics, and they are modelled through Electricity Consumption (EC) and Heating and Cooling Degree-Days (HDD and CDD, respectively). The electricity demand is interrelated with ambient conditions, i.e., heating and cooling requirements, here accessed through technical indexes based on weather conditions, HDD and CDD variables, which describe the requirements of the energy demand for heating and cooling (air conditioning) of buildings. They are derived from meteorological observations of the air temperature and interpolated in regular networks with a resolution of 25 km in Europe. These variables present a complementary characteristic throughout the year, i.e., they quantify the degrees Celsius required for heating in the winter months and cooling in the summer months.

The Industrial Production Index (IPI), measures changes in the volume of production of goods at short and regular intervals, relative to a period taken as a reference (2015 year). Under the assumption of stability of technical coefficients, this index also measures the trend of value added in volume. Doing so, its relation to the electricity demand also affects the electricity price.

Electricity prices also correlate with the mix of energy sources. In this sense, hydroelectric production, variability of fossil fuel prices, penetration of renewable energy sources and the extent in which electricity is imported are also included in the model through exogenous variables Hydroelectric Productivity Index (HPI), Europe Brent Spot Price (EBSP), Crude Oil Imports (COI), Special Regime Production (SRP) and electricity Import-Export Balance (IEB).

Hydroelectric generation, due to its high penetration in the Iberian electricity market, impacts considerably in the electricity prices. The Hydroelectric Productivity Index (HPI) reckons the deviation of the total amount of electric energy produced from hydro resources in a given period, in relation to that which would take place if an average hydrological regime occurred. The latter is evaluated taking into account 30 historical hydrological regimes. If HPI is higher than 1, the period under analysis is considered wet, and if HPI is lower than 1, from the hydrological point of view, it is considered dry.

The Europe Brent Spot Price is a major benchmark price for purchases of light crude oil in Europe. When aggregated with Crude Oil Imports of the Iberian countries, it allows the quantification of costs to generate electricity from fuel, such as natural gas.

In opposition to the ordinary regime production, including traditional non-renewable sources and large hydro-plants, the special regime production comprises generation from renewable sources, cogeneration, small production and production regulated by any other special regimes, such as the generation of electricity for self-consumption. The variable Renewable Special Regime Production measures the impact of this production from renewable sources in the electricity prices.

Finally, the extent to which electricity is imported or exported is evaluated through the Import-Export Balance that ultimately depends on the interconnections between Portugal, Spain and France.

It should be noted that from the variables stated above, the ones that depend on the dimension of the countries under analysis, are used in a per capita basis. Table 1 summarizes the dependent variable and independent variables that have demonstrated a high correlation with the electricity price on a monthly basis, their units and data sources.

Herein after, information of the country in the data set is given through suffixes -P and -S, for Portugal and Spain, respectively.

3. Forecasting methodology

research

Forecasting time horizons are not consensual in literature and vary in agreement with the primary objective of the analysis. Thresholds for electricity price forecasting may vary from a few minutes up to days ahead (short-term time horizons), from few days to few months ahead (medium-term time horizons) and months, quarter or even years (long-term time horizons), being the latest usually based on lead times measured in months. As previously introduced, the proposed analysis aims at forecasting electricity prices on a monthly basis ahead.

Numerous methods of forecasting electricity prices have been proposed over the last years. There are several modelling approaches, statistical models, multi-agent models, and computational intelligence techniques, which can be found in [3]. It is also noteworthy the growing use of hybrid models, combining those methodologies, as described in [18].

Table 1 Variables used for electricity price forecasting

Variable		Unit	Data source
Average monthly Electricity Price of the daily market	(EP)	euro/MWh	[9]
Electricity Consumption per capita	(EC)	kWh	[10, 11]
Heating Degree Days	(HDD)	(Nr)	[12]
Cooling Degree Days	(CDD)	(Nr)	[12]
Industrial Production Index	(IPI)	(n.a.)	[13, 14]
Hydroelectric Productivity Index	(HPI)	(n.a.)	[10, 11]
Europe Brent Spot Price FOB	(EBSP)	Dollars per Barrel	[15]
Crude Oil Imports per capita	(COI)	Metric Tons	[16, 17]
Renewable Special Regime Production per capita	(SRP)	MWh	[10, 11]
Electricity Import-Export Balance per capita	(IEB)	MWh	[10, 11]

The forecast methodology in this work uses a statistical approach, which chiefly derived from classical load forecasting. Statistical methods forecast the current price by using a mathematical combination of the previous prices and/or previous or current values of the exogenous or independent variables. The main advantage of the price forecasting based on exogenous variables is that it allows system operators to interpret some physical characteristics in the electricity price formation. In this context, and despite a large number of alternatives, Multiple Linear Regression Model (MLRM) is still among the most popular forecasting approach and is the model adopted in the current analysis.

3.1 Multiple Linear Regression Model

The MLRM is a statistical model that assumes there is a linear relationship between the dependent or predictor variables, Y , and X independent variables, the latter being exogenous, explanatory, non-stochastic and observable variables, used to explain the variation of the variable Y . A model that comprises more than one independent variable is a multiple regression between a dependent variable and a set of $n+1$ independent variables assuming a linear form and stochastic because it includes an error term [19]. The Multiple Linear Regression Model is given by Equation 1, as follows:

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + b_j X_{jt} + \dots + b_k X_{kt} + u_t$$

$$t = 1, 2, \dots, n, j = 1, 2, \dots, k$$

(1)

where b_0 is the y-intercept, b_j represents the parameters of the model, and u_t is the error term.

A casual association is not assumed between dependent and independent variables. In this sense, the dependent variable, Y , depends on a set of $n+1$ known factors and an unknown factor, being an endogenous variable, explained, stochastic or random and observable.

Typically, the linear regression model uses the following

assumptions [20]:

- The regression mode is linear, as proposed in Equation (1);
- The regressors are assumed to be fixed or non-stochastic in the sense that their values are fixed in repeated sampling;
- Given the values of the independent variables, the expected value of the error term is zero;
- The variance of each error term, given the values of independent variables, is constant or homoscedastic;
- There is no correlation between two error terms, i.e., there is no autocorrelation;
- There are no perfect linear relationships among the dependent variables, i.e., there is no multicollinearity;
- The regression model is correctly specified.

Based on the assumptions mentioned above, the most popular method for parameters estimation, the Ordinary Least Squares (OLS), provides estimators which have several desirable statistical properties, such as [21]:

- The estimators are linear, which means that they are linear functions of the dependent variable, Y ;
- The estimators are unbiased, which means that, in repeated applications of the method, on average, they are equal to their true values;
- The estimators are efficient, which means that they have minimum variance.

3.2 Measures of forecasting accuracy

The main purpose of the modelling and forecasting processes is to clearly discern the future values of the dependent variable, and the most important criterion of all is how accurately a model does this. The most familiar concept of forecasting accuracy is evaluated through the

error magnitude accuracy, e_t , which relates to the forecast error of a particular forecasting model, defined by Equation 2 [22]:

$$e_t = A_t - F_t \quad (2)$$

being A_t the actual value and F_t its forecast in the time period, t .

Although there are various measures of forecasting accuracy that can be used for forecast evaluation, in this work it is used the mean absolute percentage error (MAPE) expressed in generic percentage terms, computed by Equation 3 [20]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (3)$$

As stated previously in Section 3, electricity prices under analysis are based on a monthly temporal basis, for which data is significantly higher than zero. Under these circumstances, the MAPE measure performs satisfactorily on the forecasting accuracy evaluation.

4. Electricity price modelling and forecasting

4.1 Data sample and generic model

The modelling methodology adopted the historical data from January 2010 till December 2015, with a total of 72 observations. Data from 2016 year was used to validate the model, and data from 2017 and 2018 years were applied to produce the forecasts and to build the models, based on the previous validation from 2016 data, already working with 84 observations (January 2010 till December 2016).

The results were produced through GRETl statistical software (Gnu Regression, Econometrics and Time-series Library) for Windows.

The output model is no more than a representation of the relations between the variables at the same time set, according to Equation 1. Average monthly electricity price (EP) modelling and forecasting, for the Portuguese and Spanish markets, employs the econometric model given by Equation 4

It should be noted that models for Portuguese and Spanish markets interrelate the electricity price with explanatory variables for each country.

4.2 Electricity prices modelling for Portugal

The results obtained for Portugal with the Multiple Linear Regression Model, estimated by the application of the Ordinary Least Squares Method for 2017 year are presented in Table 2.

From the results obtained, the coefficient of determination is 0.6, which indicates that the independent variables explained about 60% of the variations that occurred in the electricity prices in Portugal. The adjusted coefficient of determination is 0.5519 which indicates that about 55% of the changes in electricity prices were explained by the variations in the independent variables. It is also possible to conclude:

- The autonomous component indicates that -36.6325 of the electricity prices for Portugal are not explained by independent variables. However, this variable does not reveal a statistically significant value;
- If the variables IPI-P, SRP-P, HDD-P, CDD-P, and HPI-P vary by one unit, the Electricity Price variable decreases. However, from those variables, only IPI-P, HDD-P, CDD-P and HPI-P reveal statistical significance;
- The variable electricity consumption per capita (EC-P) has a positive relation with the Electricity Price: if the first one varies one unit the later increases by approximately 0.1237 units. This variable is statistically significant, with a significance level of 1%;
- The variable COI-P has a positive relation with the Electricity Price: if the first one varies one unit, the Portuguese electricity price variable increases in 39.0607 units;
- From the analysis of the Electricity Import-Export Balance per capita (IEB-P), it has a direct relation with the Electricity Price, if the first one varies in one unit, the Portuguese electricity price variable increases in 22.2343 units;
- Regarding the F statistic (9.74), there is sufficient statistical evidence to verify that there are variables that assume values other than zero and, as previously mentioned, the variables included in the model explain satisfactorily the changes in Electricity Prices in Portugal;

$$EP_t = a + b_0 EC_t + b_1 HDD_t + b_2 CDD_t + b_4 IPI_t + b_5 HPI_t + b_6 EBSP_t + b_7 COI_t + b_8 SRP_t + b_9 IEB_t \quad (4)$$

$t = 1, 2, \dots, n$

Table 2 Performance measures of the estimated model for Portugal, 2017 year

	Coefficient	Error	ratio-t	p-value	Significance	VIF
a	-36.6325	25.6295	-1.429	0.1571		
IPI-P	-0.1751	0.0851	-2.057	4.32E-02	**	1.474
SRP-P	-17.2485	23.0701	-0.7477	4.57E-01		3.348
EC-P	0.1237	0.0331	3.734	4.00E-04	***	4.031
HDD-P	-0.07268	2.06E-02	-3.527	7.00E-04	***	7.116
CDD-P	-0.1085	4.87E-02	-2.230	2.88E-02	**	2.773
HPI-P	-11.6943	2.2848	-5.118	2.36E-06	***	1.951
EBSP-P	0.0207	0.0385	0.5364	5.93E-01		1.927
COI-P	39.0607	24.7522	1.578	1.19E-01		1.76
IEB-P	22.2343	18.4084	1.208	2.31E-01		3.375
Mean var.dependent		44.4036	D.S. var. dependent		10.5654	
White Test (TR²)		44.82	Durbin-Watson		1.086	
R²		0.6	R² adjusted		0.5519	
F (9.74)		12.36	p-value (F)		1.05E-11	

Notes: *, Significance of 10%; **, Significance of 5%; ***, Significance of 1%.

- From the analysis of the violation of the basic hypotheses of the model, in terms of multicollinearity and based on the values of the Variance Inflation Factor (VIF), there is no violation of the basic hypothesis of multicollinearity, since the VIF values, for all variables, are lower than 10. It can be concluded that there is no dependence on explanatory variables;
- Regarding the residue analysis, normality was evaluated using the Kolmogorov-Smirnov test made through the statistic test $\chi^2 = 0.6701$, with p-value=0.7351, which means that this model follows a normal distribution at a significance level of 1%, so this hypothesis is not violated. The mean is equal to $\mu = 5.0753E-15$, i.e., approximately zero, from which the zero-mean hypothesis is also not violated $E(\mu)=0$;
- For homoscedasticity, a constant variance of the error term was verified by the White test for heteroskedasticity and the test statistic $TR^2 = 45.6799$ with test value ($\chi^2(54) > 45.6799$) = 0.7825. Because the test value is higher than 10%, it can be concluded that there is no violation of homoscedasticity, i.e., the variance is constant from observation for observation. There is no loss of the characteristics of OLS estimators, since they remain BLUE;
- The Durbin-Watson statistic=1.0863 lies in the zone of positive autocorrelation of the errors. Then, it can be concluded that there is an infringement of the independence of the error term and that this model suffers from autocorrelation of the errors. In order to correct the infraction hypothesis, the Cochrane-Orcutt test was applied. Accordingly, the following statistic was obtained: Durbin-Watson=2.0855, which is now in the zone of

independence of the errors.

In order to be able to model and predict electricity prices for 2018 year, it was necessary to create a trend line from the price of electricity for Portugal and create 12 dummies (dm) or periodic auxiliary variables that represent each of the months of the year of 2018. These auxiliary variables were created as assistance to the model, due to the absence of data from the independent variables referring to the year 2018, from September 2018. In addition, to remove the trend component, a variable has been eliminated (for instance, dm2), and the least squares method was applied. The performance measures of the model are presented in Table 3.

It is also necessary to verify that the obtained model for 2018 does not violate the infractions in order to be able to validate it. Based on the performance measures of the current model and regarding the violation of the model hypotheses, it can be concluded that:

- All auxiliary variables are statistically significant at a significance level of 1%;
- There is no violation of the basic hypothesis of multicollinearity, considering the low values of the Inflation Factor of the variance;
- Constant variance of the error term, by White test for heteroscedasticity, since the value of evidence is higher than 10%, there is no violation of homoscedasticity;
- The Durbin Watson statistic=1.2243 was found in the zone of positive autocorrelation of the errors, meaning that there is an infraction to the independence of the error term. To overcome the previously verified infraction, the Cochrane-Orcutt test was

Table 3 Performance measures of the model with periodic auxiliary variables for the Portuguese market, 2018 year

	Coefficient	Error	ratio-t	p-value	Significance
Time	0.2013	0.0458	4.395	3.21E-05	***
dm1	36.4268	5.1922	7.016	5.42E-10	***
dm3	25.998	5.2276	4.973	3.44E-06	***
dm4	24.8168	5.2458	4.731	8.94E-06	***
dm5	32.2443	5.26	6.125	2.81E-08	***
dm6	37.9767	5.28	7.188	2.49E-10	***
dm7	39.353	5.30245	7.422	8.60E-11	***
dm8	38.6342	5.32199	7.259	1.80E-10	***
dm9	40.9517	5.34185	7.666	2.81E-11	***
dm10	41.5504	5.36202	7.749	1.92E-11	***
dm11	37.9729	5.38251	7.055	4.55E-10	***
dm12	41.8	5.40	7.732	2.08E-11	***
Mean var. dependent		<i>45.4093</i>	D.P. var. dependent		<i>10.58904</i>
White Test (TR²)		<i>62.5709</i>	Durbin-Watson		<i>1.2243</i>
R²		<i>0.9256</i>	R² adjusted		<i>0.4561</i>
F (12.84)		<i>87.1428</i>	value P(F)		<i>4.14E-42</i>

Notes: *, Significance of 10%; **, Significance of 5%; ***, Significance of 1%.

applied. Accordingly, the following statistic was obtained by Durbin-Watson=2.02538 which translates in independence of the errors.

4.3 Electricity prices modelling for Spain

Following the same methodology described in previous section, the model obtained for Spanish market, in 2017 year (presented in Table 4), presents a coefficient of determination of 0.5826 and indicates that the variables EC-S, HDD-S, CDD-S, HPI-S, IPI-S, EBSP-S, SRP-S, COI-S, IEB-S explain 58.3% changes of electricity prices in Spain during 2017 year. The adjusted coefficient of determination is 0.53, which indicates that about 53% of the changes in electricity prices in Spain are explained by the independent variables.

Based on the results obtained and presented in the table above, it can be concluded that:

- The autonomous component shows that -0.3407 of electricity prices in the Spanish market are not explained by the independent variables. This variable is not a statistically significant variable;
- Variables IPI-S, SRP-S, HDD-S, CDD-S, HPI-S, and IEB-S have an inverse relationship with the Electricity Price. From those, only variables HDD-S and HPI-S are statistically significant;
- EC-S, EBPS-S, and COI-S have a positive relation with the Electricity Price but only the EC-S reveals to be statistically significant, at a significance level of 5%;
- Regarding *F* statistic $F(9.74)=11.47413$, with a test value lower than 1%, there is sufficient statistical

evidence that there are variables that assume values different from zero and as previously mentioned, the variables included in the model explain in a satisfactory way the variations occurred in the electricity prices in Spain;

- The analysis of the infraction to the basic hypotheses of the model, considering the VIF, it is verified that there is no infringement of the basic hypothesis of multicollinearity (all variables present VIF lower than 10). There is no correlation between the explanatory variables;
- The test of normality of the residue performed through the statistic test $\chi^2 = 0.767096$, with test value = 0.68144, means that this model follows a normal distribution at a significance level of 1%, so this hypothesis not violated. The mean value is approximately zero, so the zero-mean hypothesis is also not violated $E(\mu) = 0$;
- Constant variance of the error term, through the White test for heteroskedasticity and the test statistic, is higher than 10%, i.e., there is no violation of homoscedasticity.

Regarding the model and prediction of the electricity prices for the Spanish market, for 2018 year, following the same methodology stated in the previous section, the model performance measures are presented in Table 5.

From the statistical tables proposed by Durbin and Watson [23], for 9 independent variables the lower bound (dL) is equal to 1.4173, upper bound (dU) equals 1.8876, 4-dU is equal to 2.1124 and finally 4-dL is equal to 2.5827. It was obtained the following Durbin-Watson

Table 4 Performance measures of the estimated model for Spain, 2017 year

	Coefficient	Error	ratio-t	p-value	Significance	VIF
a	-0.3407	26.6312	-0.01279	9.90E-01		
IPI-S	-0.1539	0.1159	-1.327	1.89E-01		1.744
SRP-S	-21.2740	17.464	-1.218	2.27E-01		2.378
EC-S	0.07358	0.0309	2.381	1.98E-02	**	4.189
HDD-S	-0.03614	1.59E-02	-2.270	2.61E-02	**	7.297
CDD-S	-0.0856	5.70E-02	-1.501	1.38E-01		6.597
HPI-S	-12.8954	1.7658	-7.303	2.67E-10	***	1.187
EBSP-S	0.04761	0.04266	1.117	2.68E-01		2.429
COI-S	59.5177	41.24	1.443	1.53E-01		1.925
IEB-S	-5.75140	46.5262	-0.1236	9.02E-01		2.33
Mean var. dependent		44.3088	D.S. var. dependent		10.1817	
White Test (TR²)		51.4569	Durbin-Watson		0.8656	
R²		0.58255	R² adjusted		0.53178	
F (9.74)		11.47413	p-value (F)		4.81E-11	

Note: *, Significance of 10%; **, Significance of 5%; ***, Significance of 1%.

Table 5 Performance measures of the model with periodic auxiliary variables for the Spanish market, 2018 year

	Coefficient	Error	ratio-t	p-value	Significance
Time	0.2073	0.0449	4.618	1.38E-05	**
dm1	36.718	5.08961	7.214	2.21E-10	***
dm3	25.7472	5.12433	5.025	2.80E-06	***
dm4	24.4437	5.14219	4.75	8.18E-06	***
dm5	31.9601	5.16	6.193	2.09E-08	***
dm6	37.4616	5.18	7.234	2.02E-10	***
dm7	38.8555	5.1977	7.476	6.73E-11	***
dm8	38.1383	5.2169	7.311	1.43E-10	***
dm9	40.2047	5.2363	7.678	2.66E-11	***
dm10	40.7012	5.2561	7.744	1.97E-11	***
dm11	37.3277	5.2762	7.075	4.16E-10	***
dm12	41.3	5.30	7.806	1.48E-11	***
Mean var. dependent		45.2971	D.P. var. dependent		10.2411
White Test (TR²)		64.1936	Durbin-Watson		1.2594
R²		0.9280	R² adjusted		-0.4959
F (12.84)		90.1925	value P(F)		1.1E-42

Note: *, Significance of 10%; **, Significance of 5%; ***, Significance of 1%.

statistic=0.865615, which lies in the zone of positive autocorrelation of the errors, meaning that there is an infringement of the independence of the term of error. Following the application of Cochrane-Orcutt test, a Durbin-Watson statistic=1.9481 is obtained, which satisfies the independence of the errors.

From the information presented in Table 5, the model for the Spanish market for 2018 year does not violate the infractions, validating it. All auxiliary variables are statistically significant with a significance level of 1%. Additionally,

- Regarding the analysis of multicollinearity, considering the VIF, it is verified that there is no violation of this hypothesis;

- White test has a test value higher than 10%, i.e., there is no violation of homoscedasticity;
- The Durbin-Watson statistic=1.2593 was obtained. This value is in the positive zone of autocorrelation of the errors, being necessary further analysis, using the test of Cochrane-Orcutt to verify if that the infraction can be solved. With the Cochrane-Orcutt test the following Durbin-Watson statistic=2.029 was obtained and, consequently, there is independence of the errors.

Table 6 Electricity prices forecast for Portugal, 2017 and 2018 years

Months	2017			2018		
	Real Price €/MWh	Forecast Price €/MWh	APE %	Real Price €/MWh	Forecast Price €/MWh	APE %
January	71.52	57.04	20.25	51.63	55.93	8.33
February	51.39	42.38	17.53	54.98	50.4	8.33
March	43.95	45.82	4.25	39.75	47.42	19.30
April	44.18	49.94	13.04	42.66	47.26	10.78
May	47.12	49.05	4.10	55.08	55.64	1.02
June	50.22	50.7	0.96	58.48	62.25	6.45
July	48.6	51.41	5.78	61.84	64.44	4.20
August	47.43	45.87	3.29	64.29	64.47	0.28
September	49.16	49.97	1.65	71.3	67.48	5.36
October	56.97	64	12.34	-	68.71	-
November	59.36	51.65	12.99	-	65.7	-
December	59.49	52.31	12.07	-	70.02	-
Mean values	52.45	50.85	9.02	55.56	59.98	7.12

Table 7 Electricity prices forecast for Spain, 2017 and 2018 years

Months	2017			2018		
	Real Price €/MWh	Forecast Price €/MWh	APE %	Real Price €/MWh	Forecast Price €/MWh	APE %
January	71.49	55.97	21.71	49.98	54.81	9.66
February	51.74	40.57	21.59	54.88	49.03	10.66
March	43.19	47.12	9.10	40.18	45.92	14.29
April	43.69	47.14	7.90	42.67	45.71	7.12
May	47.11	49.54	5.16	54.92	54.25	1.22
June	50.22	52.34	4.22	58.46	60.7	3.83
July	48.63	53.22	9.44	61.88	62.97	1.76
August	47.46	53.99	13.76	64.33	63.06	1.97
September	49.15	53.8	9.46	71.27	65.88	7.56
October	56.77	52.89	6.83	-	67.07	-
November	59.19	52.15	11.89	-	64.32	-
December	57.94	44.48	23.23	-	68.91	-
Mean values	52.22	50.27	12.02	55.40	58.55	6.45

4.4 Forecasting results for Portugal and Spain

This section presents the forecasts for the Electricity Price, for each of the country' markets under analysis, for 2017 and 2018 years, based on the models created and described in the previous sections.

To evaluate the accuracy of the prediction, it will be used the Absolute Percent Error (APE) and Mean Absolute Percent Error (MAPE). The assumed confidence interval to produce forecasts is 95%. The results obtained for the two models selected, with the previous described methodologies, and for the respective statistical measures/indicators are presented in Table 6 and Table 7, for Portugal and Spain, respectively.

Regarding the Portuguese market (Table 6) and 2017 year, it can be observed that the difference between the actual and expected annual averages is € 1.6 and the MAPE is 9.02%. Forecasts for 2017 follow the behaviour of real historical prices. For 2018, and considering the known prices, that is, between January 2018 and September 2018, it is notorious that the forecast follows the same behaviour. The MAPE, evaluated for 9 months, equals 7.12%.

From the analysis of the data of average monthly electricity price for the Portuguese Market, considering the period of analysis from January 2017 to September 2018, it is verified that this indicates maximum values in the winter months, where variables such as EC-P and HDD-P are higher which may justify the increase in electricity prices. Extrapolating this analysis to the

remaining periods, it is possible verify that the electricity prices register lower values for the summer months, where the EC-P is smaller. The minimum values of the electricity Prices are registered in March and April, for both years under analysis. This decrease in price is justified when the months have a very high HPI-P, from which higher-cost energy sources can be withdrawn from service, contributing to the decrease of the Electricity Price.

Regarding electricity prices for Spain (Table 7), it is observed that the forecast values for 2017 are higher, by € 1.95. The MAPE obtained for 2017 was 12.02%, higher than its counterpart from Portuguese market forecast and for the year 2018, about 6.45%. Analysing the year of 2017, it can be verified that the predictions follow the same behaviour of the original series, which allows trusting the model. With reference to the forecast of the average monthly electricity price for the Spanish market, maximum values are also found in winter months, where variables such as EC-S and HDD-S are higher. Similar to the results obtained for Portugal, it can be verified that the electricity prices register low values in summer months, when the EC-S is lower.

5. Conclusions

This paper presented a statistical model with explanatory variables for long-term electricity price forecasting in the Iberian electricity market. The establishment of such a reference model presents itself as an opportunity to interpret their components, intending to understand the complexity associated with price forecasting.

Regarding the Portuguese market, variables reflecting the production of goods (Industrial Production Index), ambient conditions (Heating and Cooling Degree Days), hydroelectric potential (Hydroelectric Productivity Index) and demand (Electricity Consumption per capita) are statistically significant. As far as the Spanish market is concerned, only the variables Hydroelectric Productivity Index, Heating Degree Days and Electricity Consumption per capita, are statistically significant. Therefore, it is possible to state that the electricity price in the Iberian electricity market is mainly interrelated with the inputs demand, ambient weather conditions and generation capacity.

From the analysis of the performance of the developed models, the model for the Portuguese electricity market for the year 2017, presents better results than the model applied for the Spanish electricity market. Regarding the forecast models for the year 2018, the model developed for Spain gives the best performance and the lowest MAPE.

The developed modelling suggests that factors with higher impact in the Portuguese electricity market may not be the same factors which influence the neighbouring Spanish market, even though they share to the same energy market.

The quality of the estimated models validates the use of statistical or causal methods, such as the Multiple Linear Regression Model, as a plausible strategy to obtain causal forecasts of electric energy prices in medium and long-term electricity price forecasting.

Future work will explore other approaches beside regression models to compare forecasting performances, strengths and weaknesses of different statistical techniques. A comparison with autoregressive-type time series models, relating the electricity price to its own past, and also a hybrid approach, adding the effect of the most notable exogenous variables should also be performed.

References

- [1] J. Borges, "MIBEL and the market splitting," in *5th International Conference on the European Electricity Market*, Lisboa, Portugal, 2008, pp. 1–5.
- [2] R. da Costa, "Previsão probabilística dos preços de energia elétrica do mercado ibérico de eletricidade," M.S. thesis, Faculdade de Economia, Univ. do Porto, Porto, Portugal, 2015.
- [3] R. Weron. (2014, October) Electricity price forecasting: A review of the state-of-the-art with a look into the future. [Online]. Available: <https://doi.org/10.1016/j.ijforecast.2014.08.008>
- [4] J. Ramos and *et al.*, "Structural analysis, modelling and forecasting of electricity prices in the iberian electricity market," Congreso Iberoamericano de Ciudades Inteligentes(ICSC-CITIES 2018), Soria, Esp., 2018.
- [5] S. J. Koopman, M. Ooms, and M. A. Carnero. (2007, March) Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. [Online]. Available: <http://doi.org/10.1198/016214506000001022>
- [6] C. R. Knittel and M. R. Roberts. (2005, September) An empirical examination of restructured electricity prices. [Online]. Available: <https://doi.org/10.1016/j.eneco.2004.11.005>
- [7] R. Weron and A. Misiorek. (2008, October) Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. [Online]. Available: <https://doi.org/10.1016/j.ijforecast.2008.08.004>
- [8] A. Gianfreda and L. Grossi. (2012, November) Forecasting italian electricity zonal prices with exogenous variables. [Online]. Available: <https://doi.org/10.1016/j.eneco.2012.06.024>
- [9] (2018) Price report. OMIE. [Online]. Available: <https://bit.ly/300IAvR>
- [10] (2018) Estadística mensual-sen. Redes Energéticas Nacionais. Accessed Feb. 10, 2018. [Online]. Available: <https://bit.ly/2KaiFY0>
- [11] (2018) National statistical series. Red Eléctrica de España. Accessed May. 05, 2018. [Online]. Available: <https://bit.ly/2CQ7Wy5>
- [12] (2018) Cooling and heating degree days by country-monthly data. EUROSTAT. Accessed May. 05, 2018. [Online]. Available: <https://bit.ly/2W9LBq0>
- [13] (2015) Índice de produção industrial. Instituto Nacional de Estatística. Accessed Jun. 12, 2018. [Online]. Available: <https://bit.ly/2EBNhx3>
- [14] (2015) Índice de producción industrial. Instituto Nacional de

- Estatística. Accessed May. 05, 2018. [Online]. Available: <https://bit.ly/2McGrFW>
- [15] Energy information administration. U.S. Energy Information Administration. Accessed Sep. 20, 2018. [Online]. Available: <https://bit.ly/2wqlsnf>
- [16] Portugal crude oil imports. YCharts. Accessed Sep. 20, 2018. [Online]. Available: <https://bit.ly/2VSd8Z9>
- [17] Spain crude oil imports. YCharts. Accessed Sep. 20, 2018. [Online]. Available: <https://bit.ly/2XaSfcQ>
- [18] H. V. Haghi and S. M. Tafreshi, "Modeling and forecasting of energy prices using non-stationary markov models versus stationary hybrid models including a survey of all methods," in *IEEE Canada Electrical Power Conference*, Montreal, Que., Canada, 2007, pp. 429–434.
- [19] C. Chaves, E. Maciel, P. Guimarães, and J. C. Ribeiro, *Instrumentos Estatísticos de Apoio à Economia*. New York, USA: McGraw-Hill, 2000.
- [20] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, vol. 48, no. 4, pp. 817–838, May 1980.
- [21] M. Verbeek, *A Guide to Modern Econometrics*, 4th ed. England, UK: John Wiley & Sons, 2012.
- [22] D. Gujarati and D. Porter, *Essentials of Econometrics*, 4th ed. New York, USA: McGraw Hill, 2010.
- [23] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression III," *Biometrika*, vol. 38, no. 1, pp. 159–177, Jun. 1951.